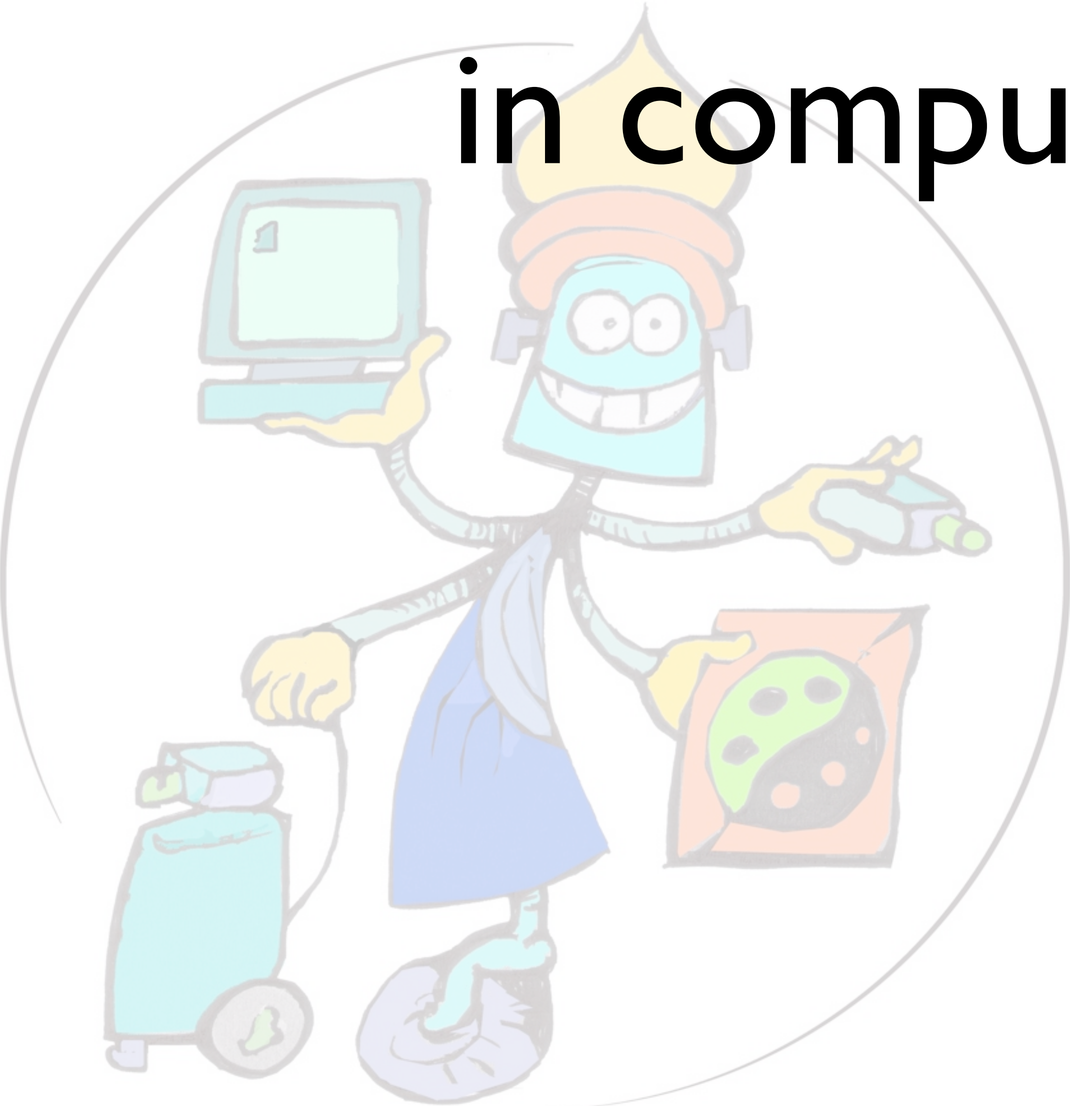
what we *did* and *do* in computer vision



francesca odone

http://slipguru.disi.unige.it

HISTORICAL PERSPECTIVE

- the (late) 80's: depth estimation (for robotics), optical flow
- the 90's:
  - motion segmentation, vehicle guidance, shape analysis,...
  - ...then machine learning comes into the picture
- 2000 and beyond:
  - 3D object recognition in videos, face detection and recognition, image & video understanding...
  - ...and back to geometry (to land on Mars)

SUMMARY

- Computer Vision at Slipguru today
  - the general approach
  - example problems:
    - (we are good at) finding and recognizing **faces** and **3D objects** in cluttered environments
    - (we are getting better at)learning actions and common **behaviors** in a scene

COMPUTER VISION & MACHINE LEARNING

- Image and scene understanding is a difficult problem
- the *learning from examples* paradigm has been proved effective to address such problems:
  - prior information on the problem is in the choice of appropriate datasets
  - ad hoc data representations and/or similarity measures from the computer vision literature to obtain *effective* models

KEY ELEMENTS OF OUR APPROACH

- the training set
- image representations
- the learning algorithm (back to Curzio's presentation)

FACE DETECTION

- **The problem:** finding occurrences of human faces at variable scale in images and videos
- *It is clearly a classification problem*
- A first issue is "how to represent data?"
  - Image descriptions based on finding meaningful components or *image parts* have been shown effective in a number of applications and motivated by biology

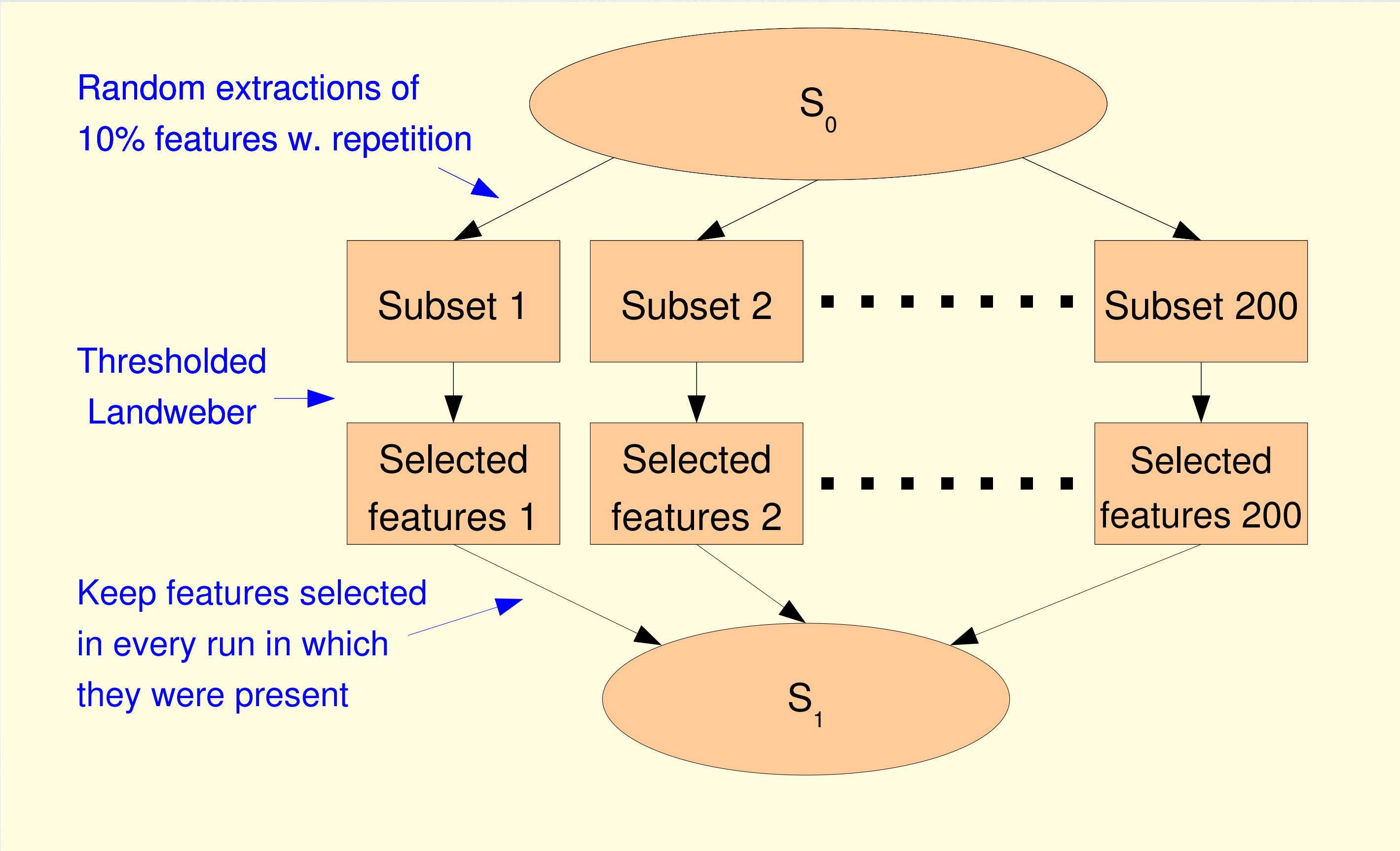
B.A. Olshauser and D. J. Field "Sparse coding with an over-complete basis set: a strategy employed by V1?" 1997.

FACE DETECTION

- Instead of extracting, or arbitrarily choosing, parts one may resort to automatic feature selection procedures
  - Start from an *overcomplete dictionary* of image measurements (e.g, Haar wavelets, rectangle features, ..)
  - Adopt a *feature selection strategy* for a data-driven selection of the most meaningful features for the problem (e.g., L1 regularization)
  - Our aim is to obtain a compact representation also suitable for real-time processing

SELECTING FACE FEATURES

$$\min_{\beta \in R^p} ||Y - \beta\Phi||^2 + \lambda||\beta||_1.$$

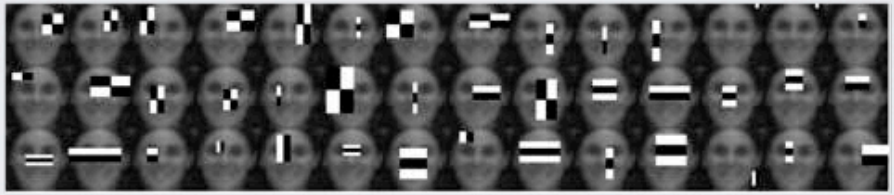


FACE AUTHENTICATION PYPELINE

Selected features at training

Test sequence

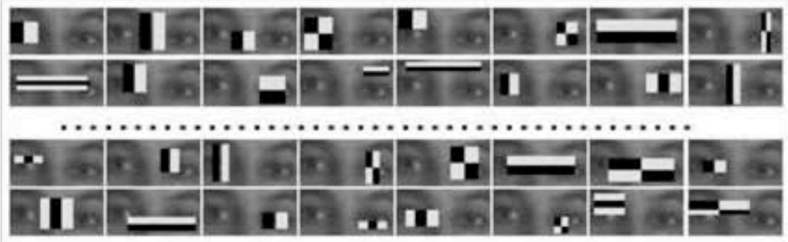
Face detection



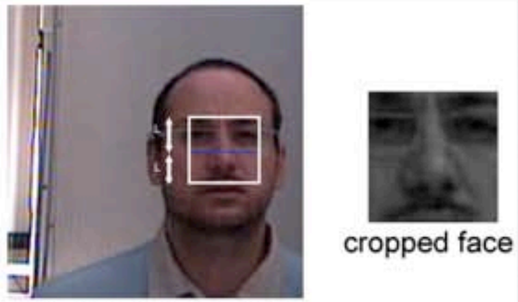
Selected rectangle features

face and eye detection

Face normalization



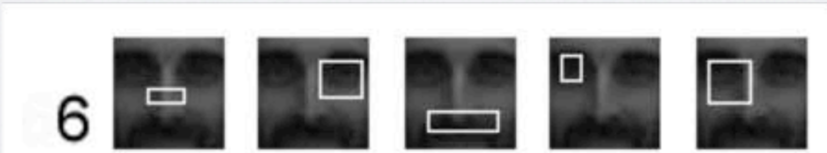
Selected rectangle features



cropped face

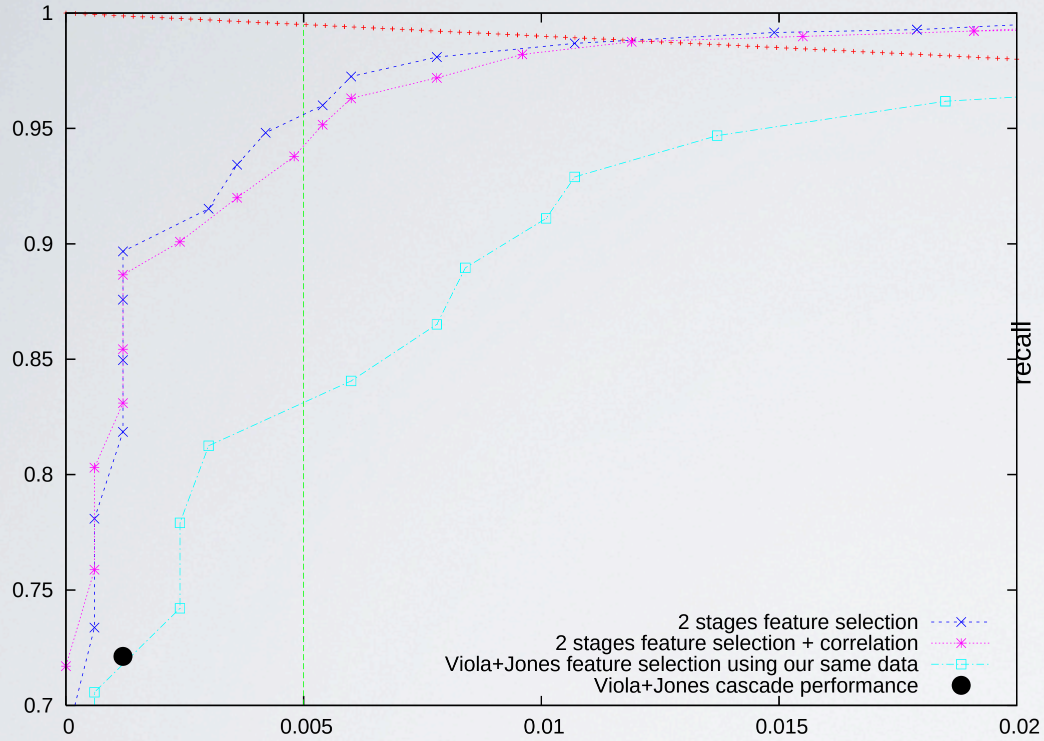
face normalization

Face authentication

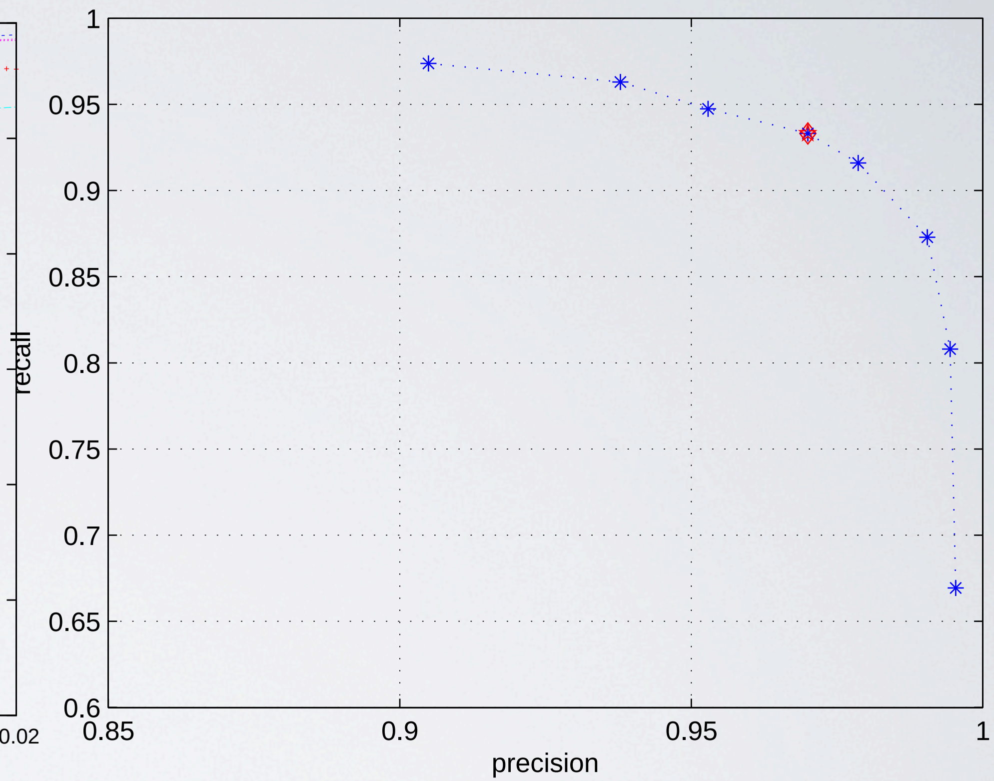


Selected LBP features

METHODS ASSESSMENT



face detection: training set of 2000+2000 esamples; test set of 2000 images



face authentication (gallery of 15 individuals;- test set of 64 indivisuals / ~ 2500 images)

A. Destrero, C. De Mol, F. Odone, A.Verri *A Regularized Framework for Feature Selection in Face Detection and Authentication*. "International Journal of Computer Vision" 83(2):164—177, 2009.

A. Destrero, C. De Mol, F. Odone, A.Verri *A sparsity-enforcing method for learning face features.* IEEE Transactions on Image Processing 18(1): 188-201, 2009.

FACE RECOGNITION FOR VISUALLY IMPAIRED USERS

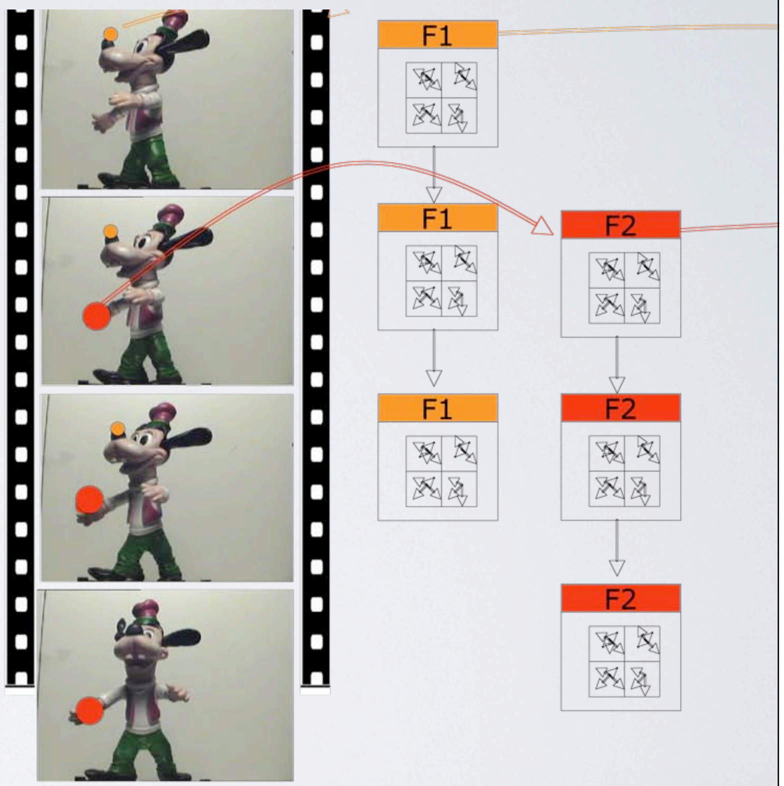
Funded by Fondazione Carige (2010)
collaboration with Istituto Chiossone

FACE AUTHENTICATION PROTOTYPE AT DISI

Funded by PSTL (2008) - Developed with
Imavis - Running since january 2008

3D OBJECT RECOGNITION WITH SPACE-TIME LOCAL FEATURES

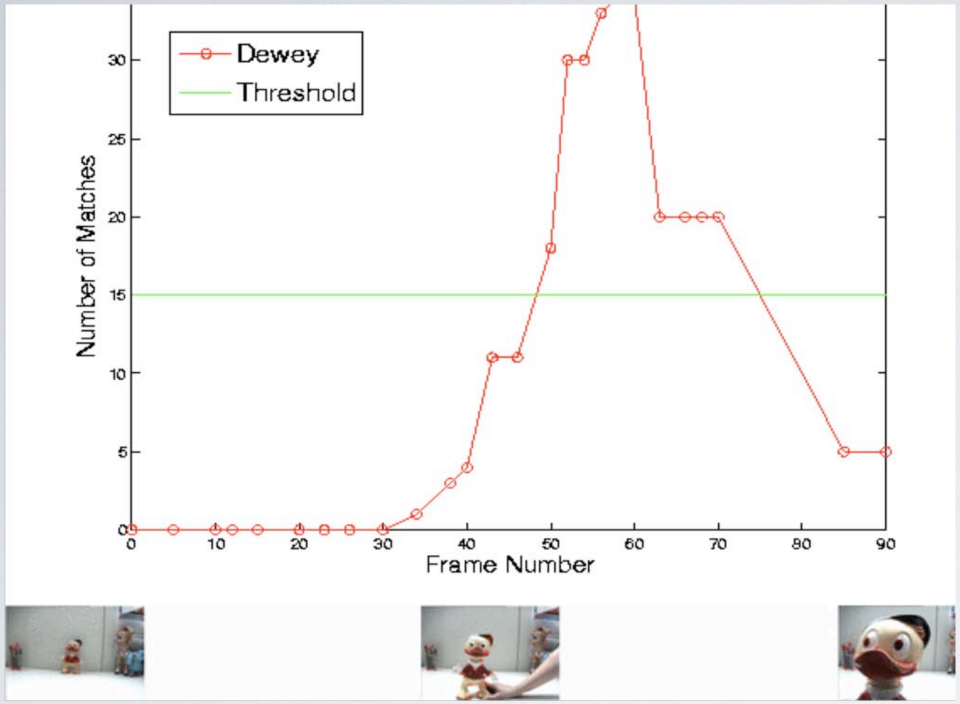
The problem: online recognition of 3D objects in cluttered environments.

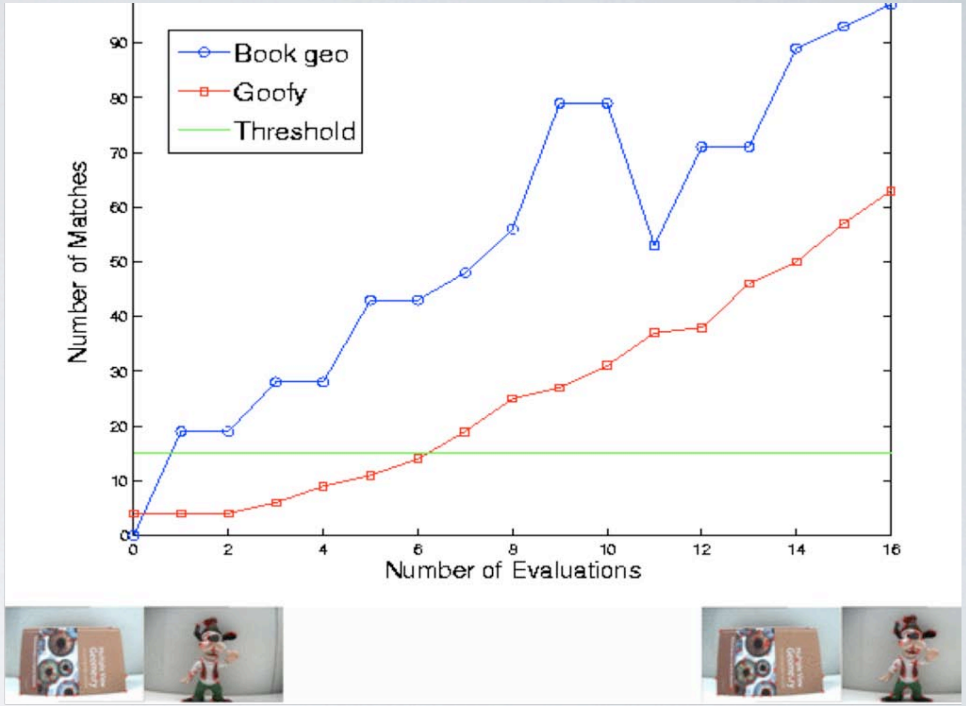


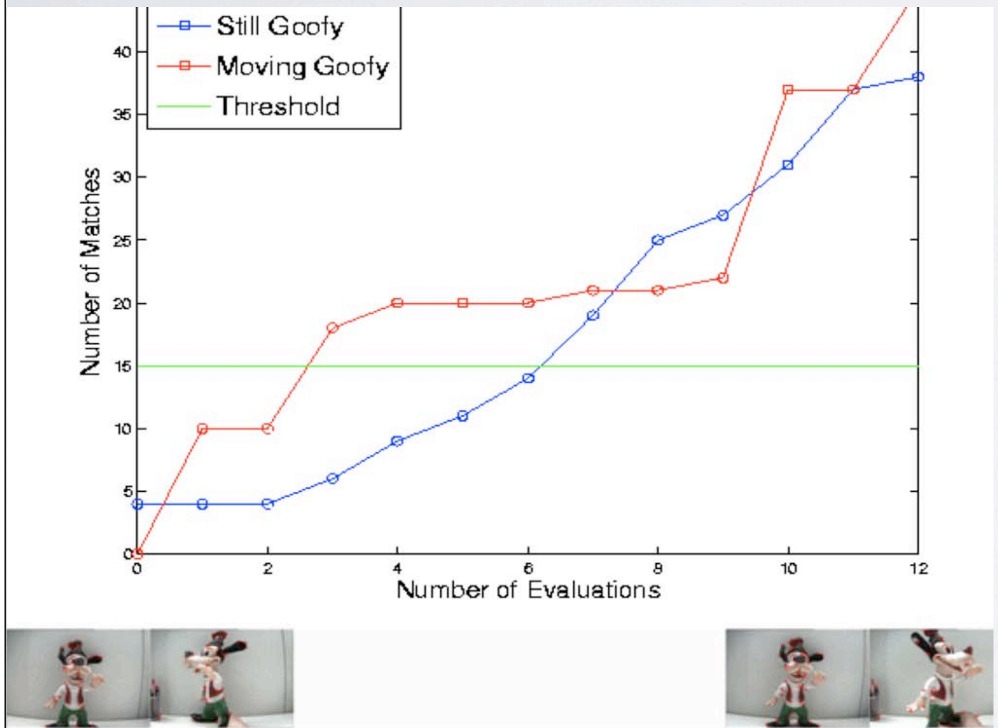
Data representation: we learn a *codebook* model per each object based on local features tolerant to view-point changes (easy to track)

Online recognition on a temporal window with a matching procedure based on space-time consistency

N. Noceti, E. Delponte, F. Odone *Spatio-temporal constraints for on-line 3D object recognition in videos.* Computer Vision and Image Understanding 2009

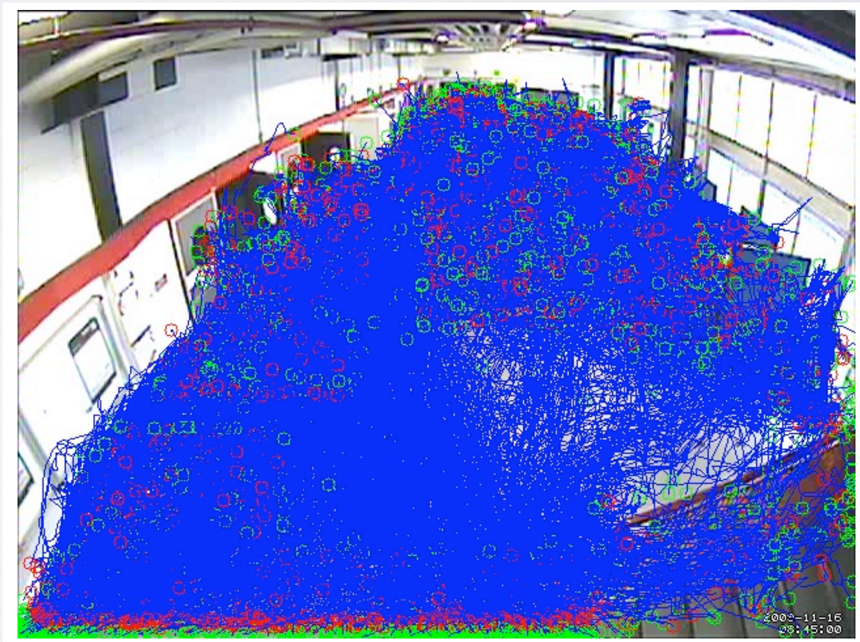

Dewey
Threshold
Number of Matches
Frame Number


Book geo
Goofy
Threshold
Number of Matches
Number of Evaluations


Still Goofy
Moving Goofy
Threshold
Number of Matches
Number of Evaluations

BEHAVIOR ANALYSIS

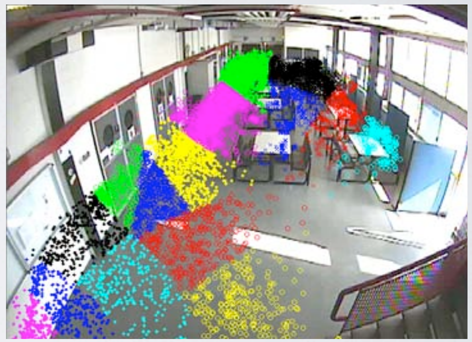
- **The problem(s):**
  - Classifying behaviors (supervised) $$Z = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$$
  - Finding recurrent patterns in the scene (unsupervised) $$Z = \{\mathbf{x}_i\}_{i=1}^{n}$$



one week events

BEHAVIOR ANALYSIS

- Data representation:
  - Multi-cue feature vector: $\mathbf{x}_i = (x_i^1, x_i^2, \ldots, x_i^{k_i})$ $x_i^t \in \mathbb{R}^d$
  - Intermediate representation based on strings

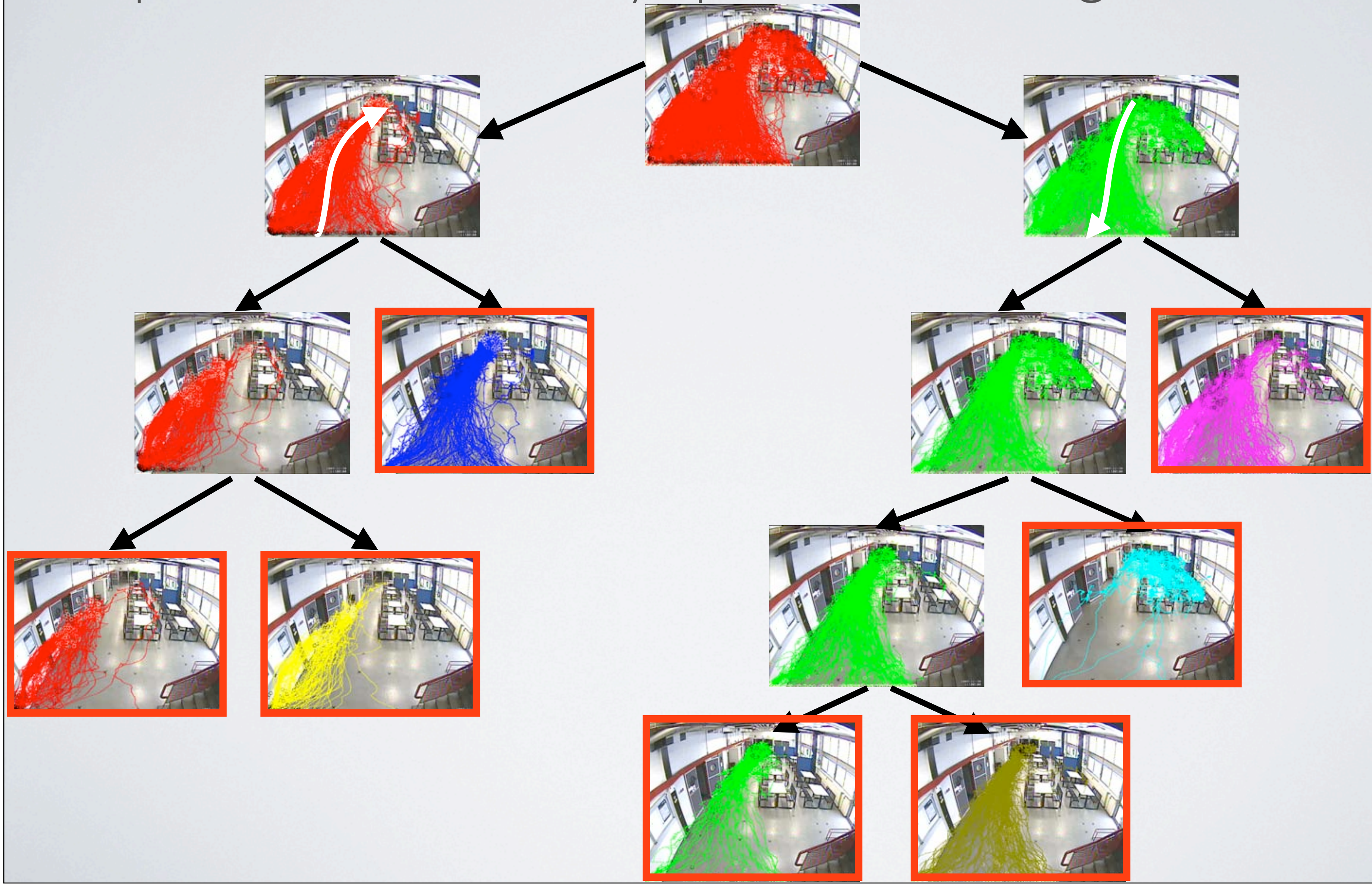


- Similarity measure (kernel)
  - p-spectrum kernel for time-sequences

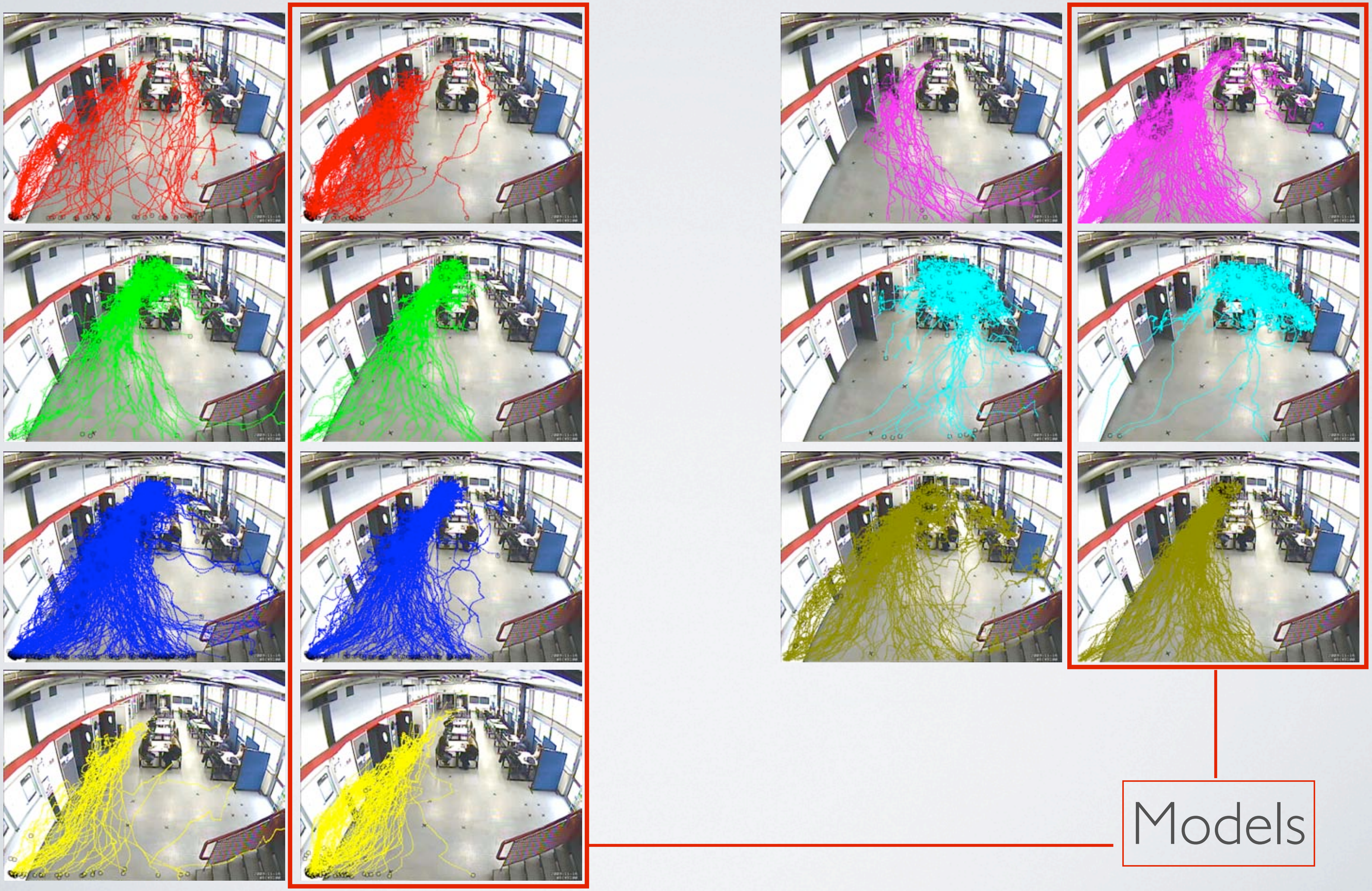
N. Noceti, M. Santoro, F. Odone. *Learning behavioral patterns of time series.* In Machine Learning from Vision-based Motion Analysis, Springer-Verlag, 2010

ESTIMATED COMMON BEHIAVIORS

Unsupervised case - 2 way spectral clustering

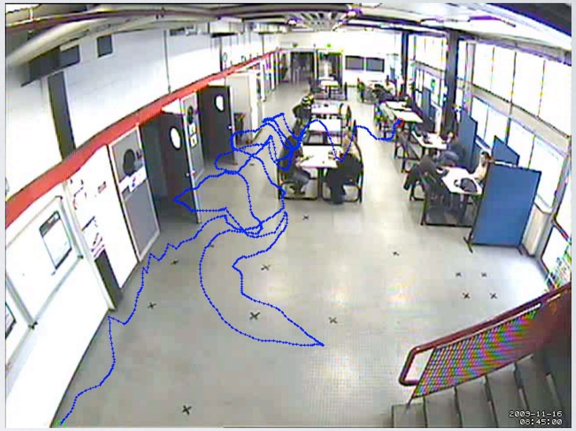
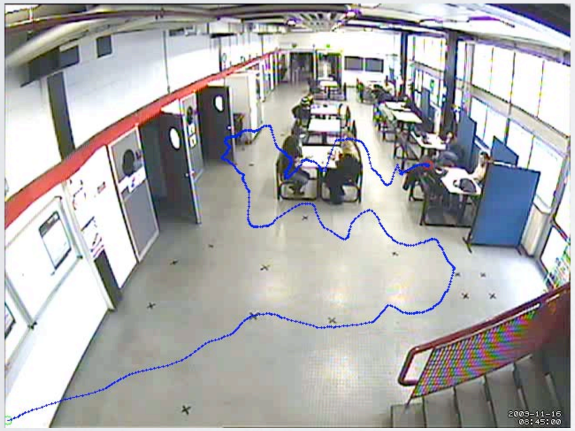
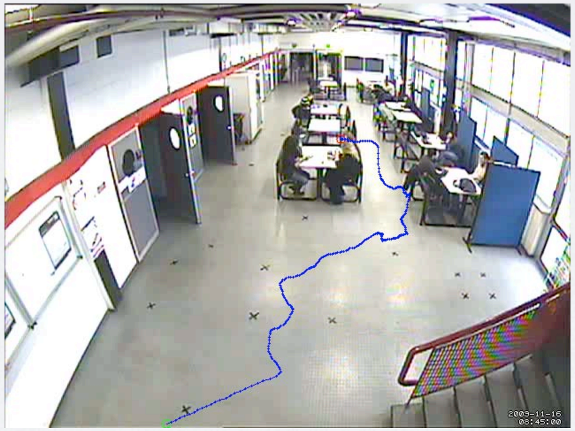


ASSOCIATION TO CLUSTERS

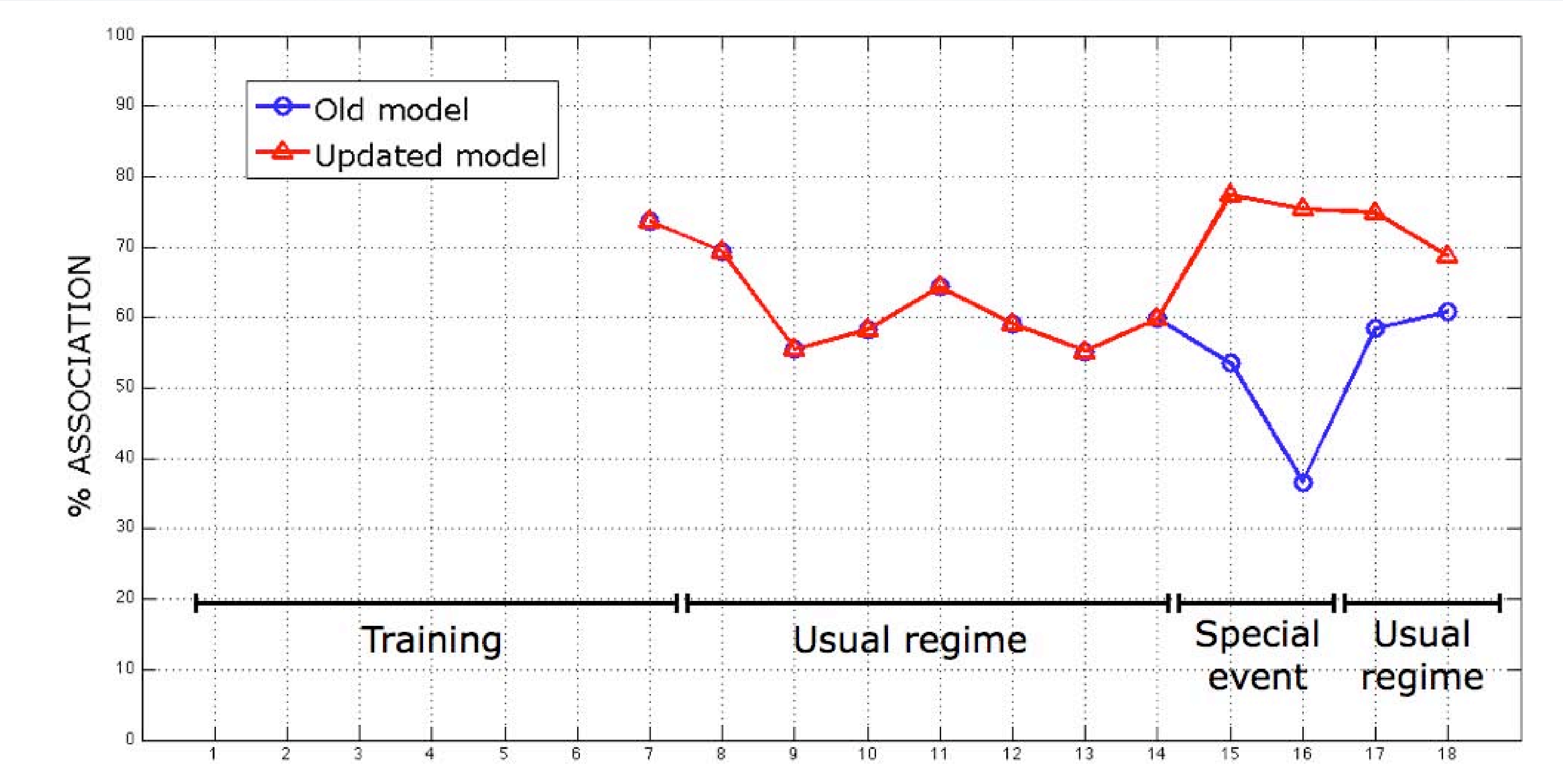


ASSOCIATION FAILURES




AS TIME GOES BY...



WHAT NOW?

- Behavior analysis:
  - CV: people groups (multi-camera coverage)
- ML for image&video classification:
  - huge amounts of data and few labels
  - the model evolves: incremental models
  - dictionary learning (see Matteo’s presentation)

REFERENCES

A. Destrero, C. De Mol, F. Odone, A.Verri *A Regularized Framework for Feature Selection in Face Detection and Authentication*. "International Journal of Computer Vision" 83(2):164—177, 2009.

A. Destrero, C. De Mol, F. Odone, A.Verri *A sparsity-enforcing method for learning face features.* IEEE Transactions on Image Processing 18(1): 188-201, 2009.

N. Noceti, E. Delponte, F. Odone *Spatio-temporal constraints for on-line 3D object recognition in videos.* Computer Vision and Image Understanding 2009

N. Noceti, M. Santoro, F. Odone. *Learning behavioral patterns of time series.* In Machine Learning from Vision-based Motion Analysis, Springer-Verlag, 2010

- PhD program “**Regularization methods for high dimensional learning**”
- Francesca Odone (UNIGE) - Lorenzo Rosasco (MIT)
- when: 5-9 july 2010
- where DISI Università di Genova